

УДК 81'276.6:615]:81-115

Р.-Ю. Т. Перхач

## КОРПУС ІНСТРУКЦІЙ ДО МЕДИЧНИХ ПРЕПАРАТІВ ЯК МЕТОД ДОСЛІДЖЕННЯ МЕДИЧНОЇ ТЕРМІНОЛОГІЇ

Перхач Р.-Ю. Т. Корпус інструкцій до медичних препаратів як метод дослідження медичної термінології.

Ця стаття висвітлює поточне і потенційне використання корпусів в лінгвістичних дослідженнях та розглядає застосування корпусного методу при аналізі медичної термінології на матеріалі українськомовних, польськомовних та німецькомовних інструкцій до медичних препаратів.

*Ключові слова:* медична термінологія, корпус текстів, інструкція до медичного препарату.

Перхач Р.-Ю. Т. Корпус инструкций к медицинским препаратам как метод исследования медицинской терминологии.

Эта статья освещает текущее и потенциальное использование корпусов в лингвистических исследованиях и рассматривает применение корпусного метода при анализе медицинской терминологии на материале украиноязычных, польскоязычных и немецкоязычных инструкций к медицинским препаратам.

*Ключевые слова:* медицинская терминология, корпус текстов, инструкция к лекарственному средству.

Perkhach R.-J. T. Corpus instructions to medicines as a research method of medical terminology.

This paper discusses the current and potential use of corpora in linguistic researches and the role of corpus-based method in analyzing the terminology of medicine in the Ukrainian, Polish and German information leaflets on medicines.

*Key words:* medical terminology, text corpus, patient information leaflet for human medicines.

У дослідженнях опрацювання медичної термінології стає необхідним використання технічних засобів, зокрема актуальними залишається дослідження в розбудові лінгвістичного корпусу інструкцій до медичних препаратів, оскільки саме інструкції до медичних препаратів містять значну кількість термінологічних одиниць і виконують роль своєрідної форми комунікації між лікарем-фахівцем та пацієнтом-неспеціалістом.

Оскільки кількість нових термінів постійно зростає, опрацювати їх

без технічних засобів є складно, тому для швидшого й ефективнішого опрацювання мовної інформації лінгвістами впроваджуються комп'ютерні технології, які дають змогу перетворювати природню мову у форму електронної інформації.

Однією з найпоширеніших форм збереження мовної інформації є лінгвістичний корпус, за допомогою якого створюється текстовий ресурс, що згодом можна використовувати багаторазово в лінгвістичних дослідженнях.

Достатньо велику кількість праць присвячено питанням корпусної лінгвістики та класифікації лінгвістичних корпусів, зокрема варто зазначити таких дослідників як, А. Баранов, М. Бейкер, О. Бугаков, О. Демська-Кульчицька, Б. Захаров, Є. Карпіловська, В. Риков, Л. Ричкова, Дж. Синклер, В. Широков [1; 3–6; 8–11].

Є чимало досліджень у галузі автоматизованого опрацювання даних, унаслідок чого з'явилося кілька значень поняття *лінгвістичний корпус* у лексичі, як-от: *електронний архів, електронні бібліотеки, колекція текстів, масив текстів, повнотекстова база даних* тощо. У зв'язку з цим виникає потреба визначити поняття *лінгвістичний корпус*. Під *лінгвістичним корпусом* текстів ми розуміємо представлені в автоматизованому вигляді, структуровані та філологічно компетентні мовні дані, призначені для розв'язання конкретних лінгвістичних завдань. У цьому разі ми дотримуємося визначення українського мовознавця О. Демської-Кульчицької, яка витлумачує *текстовий корпус* як, певним чином організоване електронне зібрання писемних і усних текстів довірливої природної мови, якому притаманні обов'язкові ознаки і яке призначене для здійснення наукового дослідження мови [3, с. 89].

Щодо класифікації лінгвістичних корпусів, то чимало дослідників у галузі корпусної лінгвістики, як-от: А. Баранов, М. Бейкер, Б. Захаров, Є. Карпіловська, Дж. Синклер, В. Риков, Л. Ричкова, О. Селіванова, виділяє кілька типів, диференційованих: за типом мовних даних (письмові, усні, змішані); за «паралельністю» (одномовні, двомовні, багатомовні); за «літературністю» (літературні, діалектні, розмовні, термінологічні, змішані); за метою створення (багатоцільові та спеціалізовані); за жанром (літературні, фольклорні, драматургічні, публіцистичні); за доступністю (вільнодоступні, комерційні та закриті); за призначенням (дослідницькі й ілюстративні); за динамічністю (динамічні і статичні); за розміткою (розмічені і нерозмічені); за характером розмітки (морфологічні, семантичні, синтаксичні, просодичні та інші); за об'ємом тексту (повнотекстові і «фрагментовані») [1, 3–6, 8–11].

Згідно з науковими розвідками О. Демської-Кульчицької існують

повнотекстові (тексти в корпусі подано повністю) та фрагментні (подані фрагменти текстів); дослідницькі (застосовують у лінгвістичних дослідженнях задля формулювання нових теорій та концепцій) ілюстративні – застосовують для підтвердження вже висловлених теоретичних положень чи гіпотез про мову; моніторингові / динамічні – забезпечують можливість відстежування змін у мові, ураховуючи аспект діакронії, а статистичні – засвідчують стан мови на певному синхронному зрізі; діакронні – репрезентують мову в понадчасовому зрізі, а синхронні репрезентують мову або тип тексту певного визначеного часового проміжку; загальномовні – репрезентують загальнонародну, національну мову, а спеціалізовані – скеровані на розв'язання часткових, галузевих, особливих, специфічних науково-дослідницьких завдань [3, с. 156–157].

Окрему ділянку корпусної лінгвістики становлять багатомовні лінгвістичні корпуси (*multilingual corpora*) та паралельні корпуси (*parallel corpora*), а також порівняльні лінгвістичні (*comparable corpora*), які є найбільш важливими для теорії та практики перекладу [10, с. 223].

Як зазначає О. Демська-Кульчицька, кількомовний корпус може містити від двох до кількох мов і за типом співвіднесення мов його далі поділяють на паралельний та порівняльний [3, с. 89].

У зв'язку з цим, на базі однакових інструкцій трьох термінологічних систем ми створили паралельний порівняльний корпус, витлумачуємо його як два або більше корпусів різних мов або різних варіантів однієї мови з однаковою структурою, побудованих за наперед визначеними параметрами. Такі корпуси містять типологічно однакові тексти цих двох і / або більше мов [3, с. 100].

Побудова порівняльного лінгвістичного корпусу фармацевтичних інструкцій проводилася поетапно. На першому етапі було сформовано джерельну базу даних із інтернет-ресурсів відповідно до анатомо-терапевтичної класифікації. Оскільки матеріал будь-якого дослідження в корпусній лінгвістиці повинен бути хронологічним та тематично однорідним, то для матеріалу нашої праці ми обрали однакові українськомовні, польськомовні та німецькомовні інструкції до медичних препаратів. Зазначені інструкції відображають аспекти всіх фармацевтичних груп, окрім ветеринарної групи медичних препаратів.

На другому етапі здійснили конвертацію файлів, а завдяки комп'ютерній програмі KWIC Concordance реалізували третій етап – токенизацію вищенаведених інструкцій. У результаті проведення токенизації було здійснено ручну лематизацію. Після завершення четвертого етапу – лематизації, відбувалася вибірка за частинами мови зі

словників та енциклопедій медичного змісту. Методом суцільної вибірки було відібрано термінологічні одиниці в українськомовних, польськомовних та німецькомовних інструкціях. Завдяки функції комп'ютерної програми KWIC Concordance створення списку словоформ за частотністю ми перейшли до кількісного аналізу.

Для систематизації та класифікації медичних термінів використано описовий метод. Цей метод сприяв поясненню особливостей терміноутворення медичних термінів. Вибрані терміни та термінологічні словосполучення диференційовано на словосполучення та словоформи, які піддавалися подальшій класифікації. Завдяки описовому методу встановлено вивчення мовних явищ на основі їх парадигматичних та синтагматичних зв'язків.

Отже, за допомогою лінгвістичного корпусу можна розв'язувати проблеми не лише з тлумаченням, але і з перекладом медичних термінів в інструкціях до медичних препаратів, якими послуговуються як лікарі, фармацевти, так і нефаківці медичної галузі, що мають потребу у використанні медичних препаратів.

Подальший аналіз інструкцій до медичних препаратів дасть змогу не лише простежити специфіку вживання медичної термінології, але й продовжити зіставний аналіз досліджуваних мов, оскільки інструкціям до медичних препаратів властиві змістовність і чіткість, що вигідно відрізняє їх від інших текстів, на основі яких здійснюють зіставний аналіз мов, зокрема слов'янських.

### Література

1. Баранов А. Н. Введение в прикладную лингвистику / А.Н. Баранов. – М., 2003. – 360 с.
2. Браунівський корпус [Електронний ресурс]. – Режим доступу : <http://khnt.aksis.uib.no/icame/manuals/brown/>.
3. Демська-Кульчицька О. Текстовий корпус : ідея іншої форми / О. Демська-Кульчицька. – К., 2011.
4. Демська-Кульчицька О. [Електронний ресурс]. – Режим доступу : <http://www.ekmair.ukma.kiev.ua/bitstream>
5. Захаров В. П. Корпусная лингвистика и лингвистические базы данных / В. П. Захаров, С. А. Коваль // НТИ: Информационные процессы и системы. – 2002. – № 7. – С. 62–69
6. Карпіловська Є. Вступ до комп'ютерної лінгвістики / Є. Карпіловська. – Донецьк, 2003.
7. Корпусна лінгвістика / В. А. Широков, О. В. Бугаков та ін. – К., 2005.
8. Рычкова Л. В. Проблема састауных аб'ектаў у корпусах славянскімоў і лінгвістычных базах дадзеных // Мовознаўства. Літэратура. Культуралогія. Фалькларыстыка. XIII Міжнародны з'езд славыстаў. Даклады беларускай дэлегацыі. – © Р.-Ю. Т. Перхач, 2015.

Мінськ, 2003. – С. 184–195.

9. Словники України on-line [Електронний ресурс]. – Режим доступу : <http://lcorep.ulif.org.ua/dictua>.

10. Baker M. Corpora in translation studies : An overview and some suggestions for future research / M. Baker. – Target 7(2). – P. 223–243.

11. Sinclair J. Corpus, Concordance, Collocation / J. Sinclair. – Oxford : Oxford University Press, 1991. – 170 p.

*Стаття надійшла до редакції 30.06.2015 р.*